

Multivariate methods for prediction of geologic sample composition with laser-induced breakdown spectroscopy

R. Anderson, R.V. Morris, S. M. Clegg, J. F. Bell III

Laser-induced breakdown spectroscopy (LIBS) uses pulses of laser light to ablate a material from the surface of a sample and produce an expanding plasma. The optical emission from the plasma produces a spectrum which can be used to classify target materials and estimate their composition. The ChemCam instrument on the Mars Science Laboratory (MSL) mission will use LIBS to rapidly analyze targets remotely, allowing more resource- and time-intensive in-situ analyses to be reserved for targets of particular interest. ChemCam will also be used to analyze samples that are not reachable by the rover's in-situ instruments. Due to these tactical and scientific roles, it is important that ChemCam-derived sample compositions are as accurate as possible.

We have compared the results of partial least squares (PLS), multilayer perceptron (MLP) artificial neural networks (ANNs), and cascade correlation (CC) ANNs to determine which technique yields better estimates of quantitative element abundances in rock and mineral samples. The number of hidden nodes in the MLP ANNs was optimized using a genetic algorithm. The influence of two data pre-processing techniques were also investigated: genetic algorithm feature selection and averaging the spectra for each training sample prior to training the PLS and ANN algorithms.

We used a ChemCam-like laboratory stand-off LIBS system to collect spectra of 30 pressed powder geostandards and a diverse suite of 196 geologic slab samples of known bulk composition. We tested the performance of PLS and ANNs on a subset of these samples, choosing to focus on silicate rocks and minerals with a loss on ignition of less than 2 percent. This resulted in a set of 22 pressed powder geostandards and 80 geologic samples. Four of the geostandards were used as a validation set and 18 were used as the training set for the algorithms.

We found that PLS typically resulted in the lowest average absolute error in its predictions, but that the optimized MLP ANN and the CC ANN often gave results comparable to PLS. Averaging the spectra for each training sample and/or using feature selection to choose a small subset of wavelengths to use for predictions gave mixed results, with degraded performance in some cases and similar or slightly improved performance in other cases. However, training time was significantly reduced for both PLS and ANN methods by implementing feature selection, making this a potentially appealing method for initial, rapid-turn-around analyses necessary for Chemcam's tactical role on MSL.

Choice of training samples has a strong influence on the accuracy of predictions. We are currently investigating the use of clustering algorithms (e.g. k-means, neural gas, etc.) to identify training sets that are spectrally similar to the unknown samples that are being predicted, and therefore result in improved predictions.